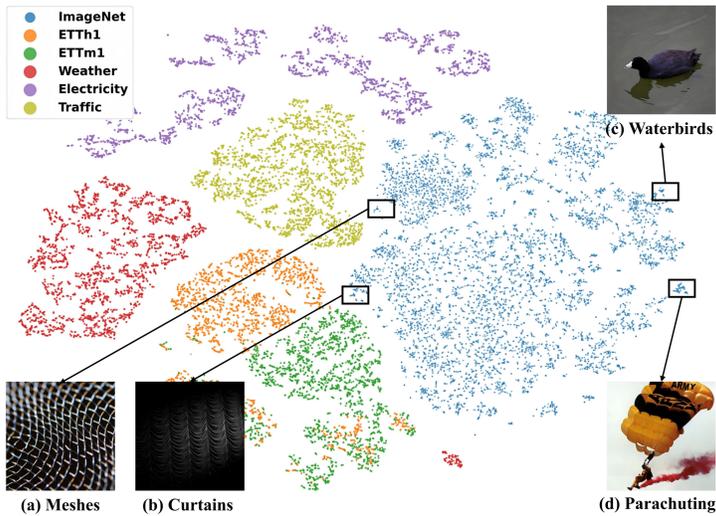


Motivation

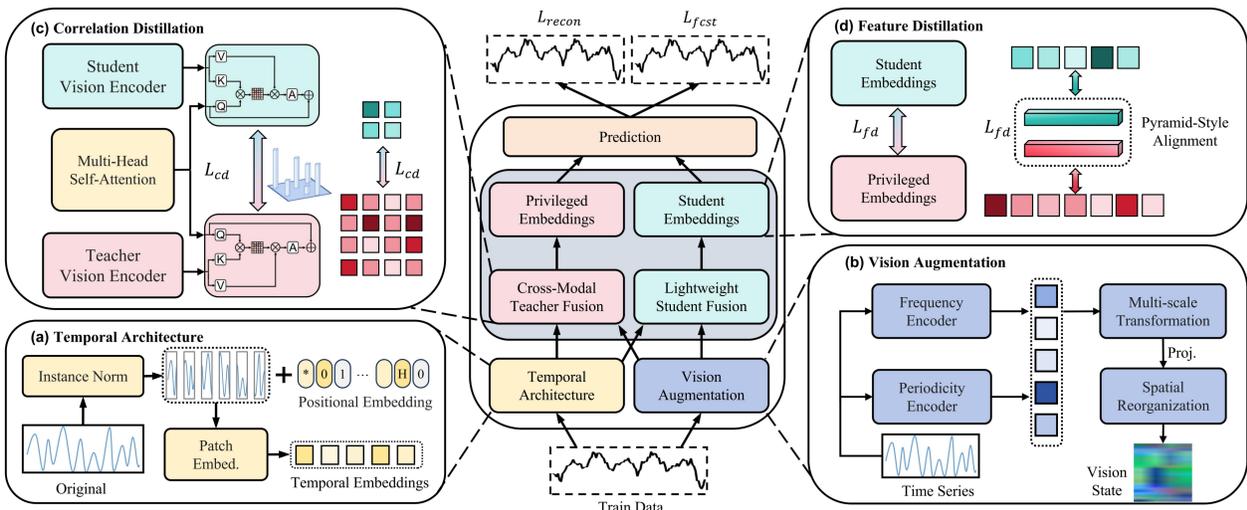
➤ Cross-Modal Analysis: Time-Series Dynamics Align with Low-Level

Visual Textures



Architecture

➤ The First Vision-Distillation Framework for Time Series Forecasting



Experiments

Methods	Ours	Only Teacher	Only Student	Time-VLM	TimeMixer++	TimeMixer	LDM4TS	TimesNet	iTransformer	DLinear	PatchTST	FEDformer	Autoformer
Metric	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE
ETT _{h1}	0.403 0.421	0.416 0.433	0.434 0.444	0.405 0.420	0.419 0.432	0.447 0.440	0.443 0.454	0.458 0.450	0.454 0.447	0.422 0.437	0.450 0.449	0.440 0.460	0.496 0.487
ETT _{h2}	0.336 0.383	0.338 0.388	0.342 0.394	0.341 0.391	0.339 0.380	0.365 0.395	0.387 0.427	0.414 0.427	0.383 0.407	0.431 0.446	0.382 0.411	0.437 0.449	0.450 0.459
ETT _{m1}	0.347 0.373	0.354 0.377	0.355 0.377	0.347 0.377	0.369 0.378	0.381 0.396	0.352 0.387	0.400 0.406	0.407 0.410	0.357 0.378	0.388 0.402	0.448 0.452	0.588 0.517
ETT _{m2}	0.245 0.307	0.252 0.313	0.258 0.317	0.248 0.311	0.269 0.320	0.275 0.323	0.333 0.380	0.291 0.333	0.288 0.332	0.267 0.333	0.293 0.336	0.305 0.349	0.327 0.371
Weather	0.224 0.259	0.229 0.268	0.230 0.269	0.224 0.263	0.226 0.263	0.226 0.263	0.229 0.277	0.259 0.287	0.258 0.278	0.248 0.300	0.258 0.280	0.309 0.360	0.338 0.382
ECL	0.162 0.259	0.168 0.267	0.170 0.270	0.172 0.272	0.165 0.253	0.182 0.273	0.199 0.299	0.192 0.304	0.178 0.270	0.166 0.263	0.204 0.294	0.214 0.327	0.227 0.338
Traffic	0.407 0.279	0.415 0.292	0.419 0.297	0.419 0.298	0.416 0.264	0.485 0.298	0.550 0.321	0.620 0.336	0.428 0.282	0.433 0.295	0.482 0.308	0.610 0.376	0.628 0.379

Table 1: Long-term forecasting results. Results are averaged over forecasting horizons $H \in \{96, 192, 336, 720\}$. Lower values indicate better performance. **Red**: best. **Blue**: second best. Full results see Appendix D.

Methods	Ours	Only Teacher	Only Student	Time-VLM	TimeMixer++	TimeMixer	LDM4TS	TimesNet	iTransformer	DLinear	PatchTST	FEDformer	Autoformer
Metric	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE
ETT _{h1} → ETT _{h2}	0.422 0.439	0.443 0.456	0.446 0.458	0.431 0.442	0.517 0.512	0.613 0.520	0.471 0.468	0.869 0.691	0.600 0.633	0.542 0.639	0.561 0.702	0.596	
ETT _{h2} → ETT _{h1}	0.344 0.390	0.356 0.402	0.357 0.403	0.356 0.402	0.379 0.391	0.402 0.433	0.452 0.460	0.479 0.465	0.428 0.438	0.605 0.538	0.415 0.431	0.466 0.475	0.488 0.499
ETT _{m1} → ETT _{m2}	0.356 0.379	0.364 0.387	0.365 0.387	0.360 0.382	0.398 0.431	0.487 0.461	0.371 0.393	0.677 0.537	0.447 0.432	0.411 0.429	0.501 0.466	0.722 0.605	0.802 0.628
ETT _{m2} → ETT _{m1}	0.253 0.313	0.261 0.321	0.262 0.322	0.263 0.323	0.291 0.351	0.311 0.367	0.336 0.373	0.320 0.353	0.295 0.338	0.316 0.368	0.296 0.343	0.463 0.488	1.342 0.930
Weather	0.227 0.262	0.230 0.268	0.231 0.269	0.233 0.274	0.241 0.271	0.242 0.281	0.229 0.276	0.279 0.301	0.272 0.290	0.241 0.283	0.242 0.279	0.284 0.324	0.300 0.342
ECL	0.181 0.283	0.206 0.310	0.209 0.312	0.188 0.291	0.168 0.271	0.187 0.277	0.172 0.275	0.323 0.392	0.202 0.288	0.180 0.280	0.180 0.273	0.346 0.427	0.431 0.478
Traffic	0.460 0.332	0.531 0.385	0.536 0.390	0.484 0.357	0.483 0.315	0.536 0.349	0.621 0.357	0.951 0.535	0.470 0.318	0.447 0.313	0.430 0.305	0.663 0.425	0.749 0.446

Table 2: Few-shot learning on 10% training data. We use the same protocol in Table 1. Full results see Appendix E.

Methods	Ours	Only Teacher	Only Student	Time-VLM	TimeMixer++	TimeMixer	LDM4TS	TimesNet	iTransformer	DLinear	PatchTST	Autoformer
Metric	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE
ETT _{h1} → ETT _{h2}	0.342 0.385	0.351 0.396	0.350 0.395	0.338 0.385	0.367 0.391	0.427 0.424	0.458 0.452	0.421 0.431	0.384 0.404	0.493 0.488	0.380 0.405	0.582 0.548
ETT _{h2} → ETT _{h1}	0.429 0.446	0.453 0.466	0.532 0.508	0.496 0.480	0.511 0.498	0.679 0.577	0.723 0.577	0.865 0.621	0.657 0.563	0.703 0.574	0.565 0.513	0.757 0.608
ETT _{m1} → ETT _{m2}	0.285 0.343	0.288 0.346	0.295 0.352	0.297 0.353	0.329 0.370	0.342 0.378	0.432 0.444	0.342 0.376	0.336 0.374	0.328 0.386	0.325 0.365	0.366 0.411
ETT _{m2} → ETT _{m1}	0.357 0.398	0.359 0.400	0.359 0.399	0.354 0.397	0.417 0.422	0.452 0.441	0.452 0.434	0.457 0.454	0.443 0.443	0.464 0.475	0.439 0.438	0.470 0.479
ETT _{m1} → ETT _{m2}	0.259 0.315	0.262 0.319	0.263 0.319	0.264 0.319	0.291 0.331	0.329 0.357	0.354 0.367	0.322 0.354	0.301 0.337	0.335 0.389	0.296 0.334	0.469 0.484
ETT _{m2} → ETT _{m1}	0.357 0.394	0.366 0.402	0.364 0.403	0.359 0.399	0.432 0.443	0.413 0.427	0.494 0.474	0.435 0.443	0.457 0.456	0.455 0.471	0.409 0.425	0.423 0.439
ETT _{h2} → ETT _{h1}	0.403 0.410	0.451 0.442	0.437 0.434	0.432 0.426	0.427 0.448	0.554 0.478	0.588 0.487	0.769 0.567	0.719 0.546	0.649 0.537	0.568 0.492	0.755 0.591

Table 3: Zero-shot learning results. We use the same protocol in Table 1. Full results see Appendix F.

Methods	Ours	Only Teacher	Only Student	TimeVLM	TimeMixer++	TimeMixer	TimesNet	iTransformer	DLinear	PatchTST	ETSformer	LightTS	FEDformer	Stationary	Autoformer	Informer
SMAPE	12.050	12.205	12.222	11.894	11.905	11.947	12.880	12.684	13.639	12.059	14.718	13.525	13.160	12.780	12.909	14.086
MASE	1.611	1.642	1.643	1.592	1.611	1.614	1.836	1.764	2.095	1.623	2.408	2.111	1.775	1.756	1.771	2.718
OWA	0.866	0.879	0.880	0.855	0.860	0.862	0.955	0.929	1.051	0.869	1.172	1.051	0.949	0.930	0.939	1.230

Table 4: Short-term time series forecasting results (Average). The forecasting horizons are in $\{6, 48\}$ and the results are weighted averaged from all datasets under different sampling intervals. Full results see Appendix G.

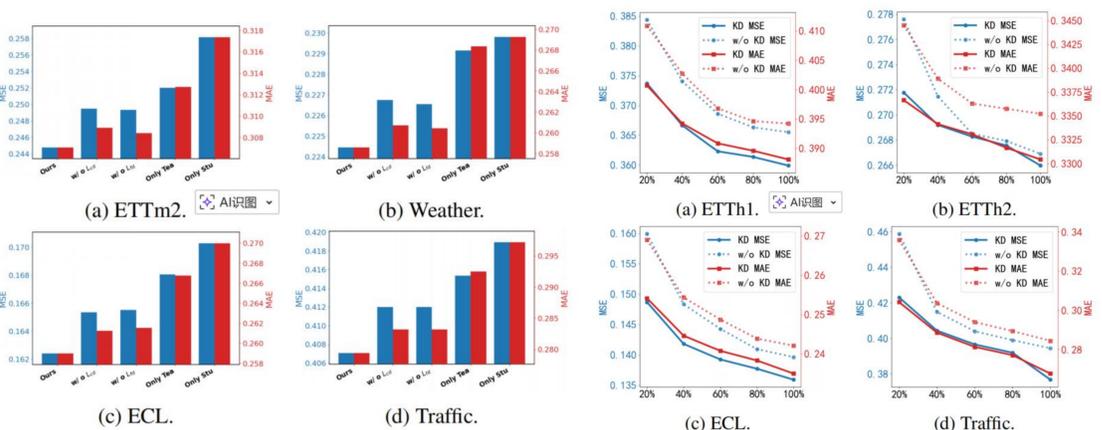


Figure 4: Ablation Experiment on Four Datasets.

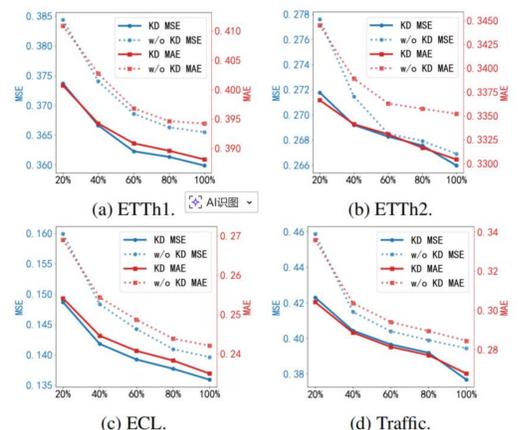
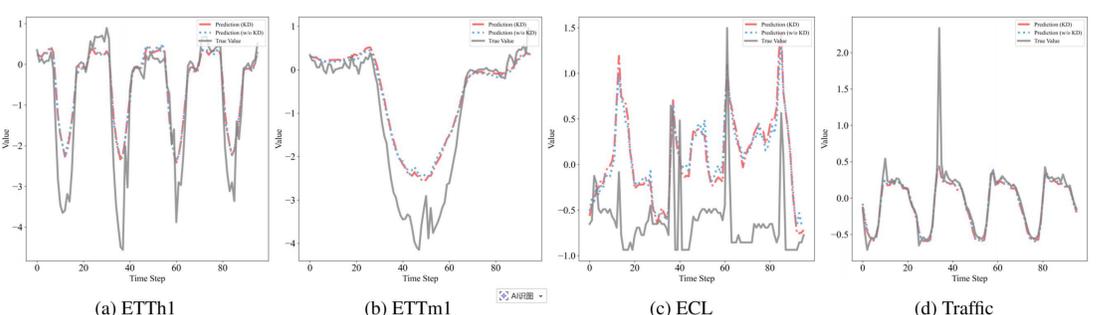


Figure 5: Effect of Different Training Data on Four Datasets.



(a) ETT_{m1}

(b) ETT_{m1}

(c) ECL

(d) Traffic

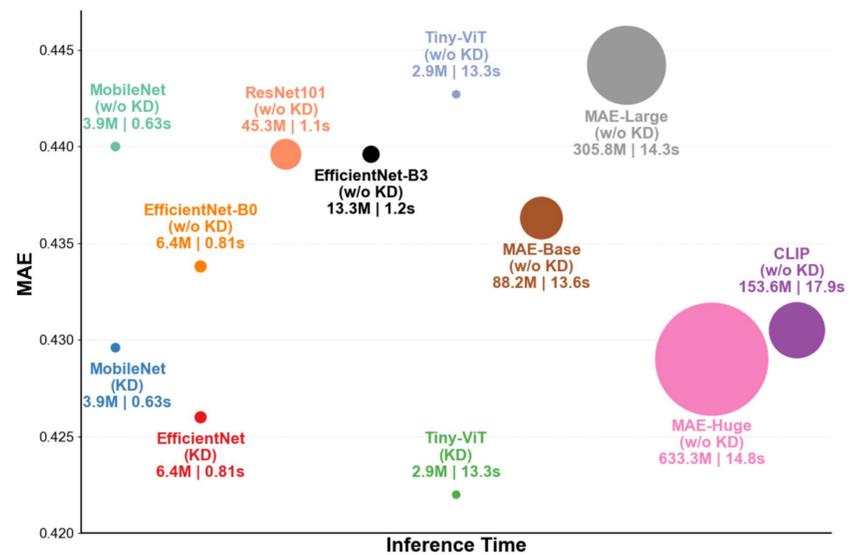


Figure 3: Model Efficiency Comparison, MAE vs Inference Time vs Parameters.



Figure 10: Visual augmentations of time series data from ETT datasets generated by our visual augmentation module. These images represent the transformed temporal patterns used as input to the vision encoders.

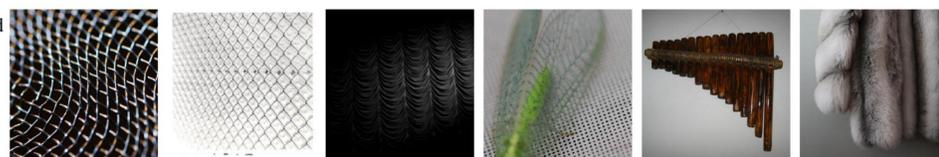


Figure 11: ImageNet images with highest feature similarity to time series representations. These nearest neighbors predominantly consist of texture-rich patterns including meshes, grids, curtains, and repetitive structures, demonstrating that time series features naturally align with low-level visual patterns.



Figure 12: ImageNet images with lowest feature similarity to time series representations. These distant samples contain semantically complex scenes with objects, animals, and human activities, confirming that high-level semantic features are largely irrelevant to time series forecasting tasks.

Paper

Code

WeChat



Feel Free To Contact Me: sisuolyu@outlook.com